# MAXIMUM LIKELIHOOD SOURCE LOCALIZATION USING THE EM ALGORITHM TO INCORPORATE PRIOR PARAMETER DISTRIBUTIONS

*Richard Perry and Kevin Buckley*

ECE Department, Villanova University, Villanova, PA 19085
buckley@ece.vill.edu, (610)-519-5658, FAX: (610)-519-4436

## ABSTRACT

In this paper we introduce a new algorithm for the estimation of source location parameters from array data given prior distributions on unknown nuisance source signal parameters. The conditional maximum-likelihood (CML) formulation is employed, and ML estimation is obtained by marginalizing over the nuisance parameters. In general, direct solution of this marginalization ML problem is intractable. We introduce an expectation-maximization (EM) algorithm solution, which is applicable to any prior distribution.

## 1. INTRODUCTION

Localization of sources of signals impinging on an array of sensors is a problem of interest in a variety of applications, including Sonar, Radar, biomedical engineering, communications and audio acoustics. A source parameter estimation approach to this problem has been well studied. Starting with received data model assumptions, numerous algorithms for location parameter estimation have been developed. For maximum likelihood (ML) and maximum a posterior (MAP) source location parameter estimation, two basic models have received considerable attention. For the conditional (or deterministic maximum-likelihood (CML) model [1,2,3,4], source signal amplitudes as well as other secondary or nuisance parameters are considered deterministic and unknown. For the stochastic maximum-likelihood (SML) model [1,3], the covariance of the source signals as well as other nuisance parameters are assumed unknown. The CML model is more generally applicable, while if valid the SML model leads to better location estimator performance.

The expectation-maximization (EM), introduced to the signal processing community by Dempster et al. [5], is a general algorithmic approach which can be used to solve ML problems. Its potential advantages include the possibility of development of algorithms which are computationally efficient for ML problems which are intractable to solve directly. For location parameter estimation, examples of the use of the EM approach include Feder and Weinstein [6] for a no secondary-parameter and a SML problem, Miller and Fuhrmann [7] for both CML and SML problems, and Ziskind and Hertz [8] for an AR source signal model problem. In none of the references cited above is prior knowledge of nuisance parameter distributions accounted for.

Compared to joint estimation on both primary and nuisance parameters, marginalization over nuisance parameters can result in improved estimator performance for the primary parameters. This is particularly true in limited-data situations and when the number of nuisance parameters grows with the amount of data. To marginalize over the nuisance parameters, a prior distribution on them must be assumed. Although accuracy of the prior distribution is important, in particular when the prior is restrictive, it has been shown that estimator performance can be improved even when noninformative (or diffuse) priors are assigned to the nuisance parameters. For source localization based on the CML model, see [4] for an illustration of this. In general, a problem with marginalization is the difficulty in solving the resulting optimization problem.

In this paper we consider, within the CML model formulation, incorporation of knowledge of prior distributions on both primary location and secondary source signal parameters. We focus on using priors on the nuisance source signal parameters to marginalize over them, and we introduce a simple EM algorithmic approach to solving the resulting optimization problem. This extends to the array based source location parameter estimation problem our recent work [9,10] on digital communication symbol estimation in the presence of secondary unknown channel parameters.

## 2. ARRAY OBSERVATION MODEL

We assume the standard narrowband observation model for $D$ far-field sources impinging on an array of $M$ sensors. Letting $\theta$ denote the vector of source location parameters to be estimated, the array observation is

$$\mathbf{x}_k = \mathbf{A}(\theta)\, \mathbf{s}_k + \mathbf{n}_k,\ k = 1, ..., N, \tag{1}$$

where $\mathbf{s}_k$ denotes the $D \times 1$ vector of signal amplitudes at time $k$, $\mathbf{A}(\theta)$ is the $M \times D$ matrix with (assumed linearly independent) array response vector columns $\mathbf{a}(\theta_i)$, for $i = 1, ..., D$. The $M \times 1$ noise vector, $\mathbf{n}_k$, is modeled as a zero-mean, temporally and spatially white Gaussian process with standard deviation $\sigma$. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$, $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, ..., \mathbf{n}_N]$, and $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N]$. Then

$$\mathbf{X} = \mathbf{A}(\theta)\, \mathbf{S} + \mathbf{N} \ . \tag{2}$$

The complex signal amplitudes in $\mathbf{S}$ are treated as a set of $ND$ unknown nuisance or secondary parameters. Our primary interest is the estimation of the location parameters $\theta$.

With this array observation model and under the CML model assumption, the probability density function of the received data (conditioned on $\mathbf{S}$ as well as $\theta$) is:

$$f(\mathbf{X}|\mathbf{S}, \theta) = \frac{1}{(\pi\sigma^2)^{NM}} \prod_{k=1}^{N} e^{-\frac{||\mathbf{x}_k - \mathbf{A}(\theta)\mathbf{s}_k||^2}{\sigma^2}}. \tag{3}$$

This research was supported by ONR under Grant N00014-98-1-0892.

The basic equations above describe the source localization problem under consideration, based only on the fact that the additive noise is spatially/temporally white and Gaussian. In the succeeding sections, additional known properties of the random signal amplitudes will be considered, starting from the most general joint pdf case and proceeding to specific examples of signal amplitude probability distribution functions.

## 3. EM AND ML SOURCE LOCATION ESTIMATION

For the MAP estimator, we must find $\theta$ to maximize $f(\theta|\mathbf{X})$. Using Bayes rule:

$$\begin{aligned} f(\theta|\mathbf{X}) &= \frac{f(\mathbf{X}|\theta)f(\theta)}{f(\mathbf{X})} \\ &= K_4 f(\mathbf{X}|\theta)f(\theta) \doteq f(\mathbf{X}|\theta)f(\theta) \ , \end{aligned} \tag{4}$$

where $K_4$ is a constant which takes into account the effect of $f(\mathbf{X})$ for normalizing the distribution.[1]

If $f(\theta)$ is unknown or assumed to be uniform, then the ML estimator which maximizes the likelihood function $f(\mathbf{X}|\theta)$ can be used instead. All of the EM algorithms derived in the subsequent sections here produce iterative solutions to the ML problem of maximizing $f(\mathbf{X}|\theta)$. But if $f(\theta)$ is known and not uniform, then this could be easily incorporated into the EM algorithms as an additive term in the cost function, thus producing a MAP estimator.

The dependence of $f(\mathbf{X}|\theta)$ on the random channel coefficients is:

$$f(\mathbf{X}|\theta) = E[f(\mathbf{X}|\mathbf{S},\theta)] = \int_{\mathbf{S}} f(\mathbf{X}|\mathbf{S},\theta)f(\mathbf{S})d\mathbf{S}, \tag{5}$$

where $f(\mathbf{X}|\mathbf{S},\theta)$ is given by (3). Evaluating this multidimensional integral and then maximizing the result with respect to $\theta$ seems to be an intractable problem in general. Yet this is exactly the problem which the EM algorithms can solve iteratively.

To maximize (5) using EM, define the auxiliary function [5]:

$$\begin{aligned} Q(\theta|\phi) &= E[\log(f(\mathbf{X},\mathbf{S}|\theta))\,|\mathbf{X},\phi] \\ &= \int_{\mathbf{S}} \log(f(\mathbf{X},\mathbf{S}|\theta))f(\mathbf{S}|\mathbf{X},\phi)d\mathbf{S}, \end{aligned} \tag{6}$$

where we desire the ML estimate of $\theta$, and $\phi$ represents the current estimate of $\theta$. In the terminology of EM, $\mathbf{X}$ is the observed data, $\mathbf{S}$ is the hidden data, and $(\mathbf{X},\mathbf{S})$ is the complete data.

The general EM algorithm for this problem is:

1. Initialize $\phi$ to an estimate of $\theta$.
2. E-step (Expectation): construct $Q(\theta|\phi)$
3. M-step (Maximization): find $\theta$ to maximize $Q(\theta|\phi)$
4. Set $\phi = \theta$ and repeat steps 2 and 3 until converged.

[1] As a matter of notation, we will subscript normalization constants such as $K_4$ throughout this paper with the equation number in which they first appear. Also, we will use the symbol $\doteq$ to mean "equivalent within an additive and/or multiplicative constant" for equations in which such constants have no effect for purposes here.

In general, to construct $Q(\theta|\phi)$, start with $f(\mathbf{X},\mathbf{S}|\theta)$:

$$f(\mathbf{X},\mathbf{S}|\theta) = f(\mathbf{X}|\mathbf{S},\theta)f(\mathbf{S}|\theta). \tag{7}$$

We can drop the $f(\mathbf{S}|\theta)$ term from this since $\mathbf{S}$ and $\theta$ are independent so $f(\mathbf{S}|\theta)$ is not a function of $\theta$ and will not affect the maximization in the M-step. $Q(\theta|\phi)$ from (6) may then be written as:

$$\begin{aligned} Q(\theta|\phi) &\doteq E[\log(f(\mathbf{X}|\mathbf{S},\theta))\,|\mathbf{X},\phi] \\ &= \int_{\mathbf{S}} \log(f(\mathbf{X}|\mathbf{S},\theta))f(\mathbf{S}|\mathbf{X},\phi)d\mathbf{S}. \end{aligned} \tag{8}$$

To evaluate $Q(\theta|\phi)$, we need $f(\mathbf{S}|\mathbf{X},\phi)$, which can be written in terms of the apriori channel coefficient density function $f(\mathbf{S})$ as:

$$f(\mathbf{S}|\mathbf{X},\phi) = \frac{f(\mathbf{X}|\mathbf{S},\phi)f(\mathbf{S}|\phi)}{f(\mathbf{X}|\phi)}, \tag{9}$$

where the denominator term $f(\mathbf{X}|\phi)$ can be treated as a constant since it simply serves to normalize the distribution, and is not a function of $\mathbf{S}$ or $\theta$, so will not affect the maximization step. Also, $f(\mathbf{S}|\phi) = f(\mathbf{S})$ since the signal amplitudes and source locations are independent. But note that given $\mathbf{X}$ and $\phi$, $f(\mathbf{S}|\mathbf{X},\phi) \neq f(\mathbf{S})$. Therefore, for $f(\mathbf{S}|\mathbf{X},\phi)$ in (8) we may use:

$$f(\mathbf{S}|\mathbf{X},\phi) \doteq f(\mathbf{X}|\mathbf{S},\phi)f(\mathbf{S}). \tag{10}$$

In the specific cases considered subsequently, we will show that the E-step reduces to computing sufficient statistics of $Q(\theta|\phi)$, which are the mean and covariance of $\mathbf{S}$ conditioned on $\mathbf{X}$ and $\phi$. This enables the use of an ML-like estimator of $\theta$ in the M-step, which employs the sufficient statistics provided in the E-step. The E-step and M-step formulas will be derived in detail in subsequent sections for various specific signal amplitude distributions.

## 4. EM ALGORITHMS FOR SOURCE LOCALIZATION

Here we show that the E-step reduces to computing sufficient statistics of $Q(\theta|\phi)$, which are the mean and covariance of $\mathbf{S}$ conditioned on $\mathbf{X}$ and $\phi$. This enables the use of an ML-like estimator of $\theta$ in the M-step, which employs the sufficient statistics provided in the E-step.

From (8) and (3), the auxiliary function $Q(\theta|\phi)$ is

$$Q(\theta|\phi) \doteq E[\,-\sum_{k=1}^{N}||\mathbf{x}_k - \mathbf{A}(\theta)\mathbf{s}_k||^2\,|\mathbf{X},\phi] \tag{11}$$

$$\begin{aligned} &= -\sum_{k=1}^{N} E[||\mathbf{x}_k - \mathbf{A}(\theta)\mathbf{s}_k||^2\,|\mathbf{X},\phi] \\ &= -\sum_{k=1}^{N}\int_{\mathbf{S}} ||\mathbf{x}_k - \mathbf{A}(\theta)\mathbf{s}_k||^2 f(\mathbf{S}|\mathbf{X},\phi)d\mathbf{S} \ . \end{aligned}$$

For each $k$, the conditional expectation over $\mathbf{S}$ is of the second order function $||\mathbf{x}_k - \mathbf{A}(\theta)\mathbf{s}_k||^2$. It can be shown that this reduces to:

$$Q(\theta|\phi) \doteq -\sum_{k=1}^{N}\left\{||\mathbf{x}_k - \mathbf{A}(\theta)\mathbf{g}_k||^2 + Tr\{\mathbf{A}(\theta)\mathbf{G}_k\mathbf{A}^H(\theta)\}\right\} \tag{12}$$

where

$$\mathbf{g}_k \triangleq E[\mathbf{s}_k|\mathbf{X},\phi] \tag{13}$$

and

$$\mathbf{G}_k \triangleq E[(\mathbf{s}_k-\mathbf{g}_k)(\mathbf{s}_k-\mathbf{g}_k)^H|\mathbf{X},\phi] \tag{14}$$

are, respectively, the mean and covariance of the $\mathbf{s}_k$ conditioned on $\mathbf{X},\phi$. A key point is that, to evaluate (12), we need only the means and covariances of the $\mathbf{s}_k$ given $\mathbf{X}$ and $\phi$ using the distribution of (9). We will derive these for specific cases of $f(\mathbf{S})$ subsequently. The noise variance $\sigma^2$ is needed in general to compute the $\mathbf{g}_k$ and $\mathbf{G}_k$.

The EM algorithm is as described generally in Section 3 above, where the E-step reduces simply to the computation of the $\mathbf{g}_k$ and the $\mathbf{G}_k$. To initialize $\phi$ in step 1, we can initialize $\mathbf{g}_k$ and $\mathbf{G}_k$, $k=1,\ldots,N$, maximize (12) to estimate $\theta$, then set $\phi=\theta$. To initialize and estimate $\mathbf{g}_k$ and $\mathbf{G}_k$ in steps 1 and 2 of the algorithm, we need to know the specific form of $f(\mathbf{S})$.

### 4.1. Temporally Independent Signal Amplitudes

If the signal amplitude vectors $\mathbf{s}_k$ are independent over time, then $f(\mathbf{S})$ can be written as:

$$f(\mathbf{S}) = \prod_{k=1}^{N} f(\mathbf{s}_k), \tag{15}$$

and $f(\mathbf{S}|\mathbf{X},\phi)$ becomes:

$$f(\mathbf{S}|\mathbf{X},\phi) = \prod_{k=1}^{N} f(\mathbf{s}_k|\mathbf{X},\phi) \quad . \tag{16}$$

From (10) and (3) we have that

$$f(\mathbf{s}_k|\mathbf{X},\phi) \doteq f(\mathbf{s}_k)\, e^{-\frac{||\mathbf{x}_k-\mathbf{A}(\phi)\mathbf{s}_k||^2}{\sigma^2}}. \tag{17}$$

For temporally independent Gaussian signal amplitudes, let $f(\mathbf{s}_k)$ be a joint Gaussian distribution with mean $\mathbf{c}_k$ and covariance $\mathbf{C}_k$:

$$f(\mathbf{s}_k) = \frac{1}{\pi^N|\mathbf{C}_k|} e^{-(\mathbf{s}_k-\mathbf{c}_k)^H\mathbf{C}_k^{-1}(\mathbf{s}_k-\mathbf{c}_k)}. \tag{18}$$

In this case (17) specifically becomes

$$\begin{aligned} f(\mathbf{s}_k|\mathbf{X},\theta) &\doteq e^{-\frac{||\mathbf{x}_k-\mathbf{A}(\phi)\mathbf{s}_k||^2}{\sigma^2}}\; e^{-(\mathbf{s}_k-\mathbf{c}_k)^H\mathbf{C}_k^{-1}(\mathbf{s}_k-\mathbf{c}_k)} \\ &\doteq e^{-(\mathbf{s}_k-\mathbf{g}_k)^H\mathbf{G}_k^{-1}(\mathbf{s}_k-\mathbf{g}_k)}, \end{aligned} \tag{19}$$

with

$$\mathbf{G}_k = \left[\frac{\mathbf{A}^H(\phi)\mathbf{A}(\phi)}{\sigma^2} + \mathbf{C}_k^{-1}\right]^{-1} \tag{20}$$

$$\mathbf{g}_k = \mathbf{G}_k\left(\frac{\mathbf{A}^H(\phi)\mathbf{x}_k}{\sigma^2} + \mathbf{C}_k^{-1}\mathbf{c}_k\right) \quad . \tag{21}$$

To initialize $\phi$ in step 1, we could, for example, initialize $\mathbf{G}_k=\mathbf{C}_k$ and $\mathbf{g}_k=\mathbf{c}_k$, $k=1,\ldots,N$.

### 4.2. Gauss-Markov Signal Amplitudes

For a Gauss-Markov time-varying signal amplitude vector, let $\alpha$ represent a complex first-order Markov factor. Let the signal amplitude vector follow a Gauss-Markov distribution such that, at time $k$:

$$\mathbf{s}_k = \alpha\mathbf{s}_{k-1} + \mathbf{v}_k, \tag{22}$$

where $\mathbf{v}_k$ is complex, white, and Gaussian with mean $\mathbf{c}$ and covariance $\mathbf{C}$.

In this case, $f(\mathbf{s}_k|\mathbf{s}_{k-1})$ is Gaussian with mean $\mathbf{c}+\alpha\mathbf{s}_{k-1}$ and covariance $\mathbf{C}$. The $Q(\theta|\phi)$ from (12) with the $\mathbf{g}_k$ and $\mathbf{G}_k$ from (13,14) still apply.

$f(\mathbf{S})$ can be written as:

$$f(\mathbf{S}) = f(\mathbf{s}_1)\prod_{k=2}^{N} f(\mathbf{s}_k|\mathbf{s}_{k-1}), \tag{23}$$

so from (10) and (3) $f(\mathbf{S}|\mathbf{X},\phi)$ becomes:

$$f(\mathbf{S}|\mathbf{X},\phi) \doteq f(\mathbf{s}_1)\prod_{k=2}^{N} f(\mathbf{s}_k|\mathbf{s}_{k-1})\prod_{k=1}^{N} e^{-\frac{||\mathbf{x}_k-\mathbf{A}_k(\phi)\mathbf{s}_k||^2}{\sigma^2}}. \tag{24}$$

To determine $f(\mathbf{s}_k|\mathbf{X},\phi)$, and thereby $\mathbf{g}_k$ and $\mathbf{G}_k$, we must integrate (24) over $\mathbf{s}_i$, $i=1,\ldots,N$, $i\neq k$. Due to space limitations, the derivation of this integration result is not shown here. A parallel derivation for a similar model and problem (a Gauss-Markov FIR intersymbol interference channel model for a digital communication problem) appears in [10]. The result is that $f(\mathbf{s}_k|\mathbf{X},\phi)$ is Gaussian, with mean $\mathbf{g}_k$ and covariance $\mathbf{G}_k$ determined in the E-step by the following recursive algorithm:

1) Initialize: for $k=N,N-1,\ldots,2$

$$\begin{aligned} \hat{\mathbf{G}}_k &= \left[\frac{\mathbf{A}_k^H(\phi)\mathbf{A}_k(\phi)}{\sigma^2} + \mathbf{C}^{-1} + |\alpha|^2\left(\mathbf{C}^{-1} - \mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\mathbf{C}^{-1}\right)\right]^{-1} \\ \hat{\mathbf{q}}_k &= \frac{\mathbf{A}_k^H(\phi)\mathbf{x}_k}{\sigma^2} + (1-\alpha^*)\mathbf{C}^{-1}\mathbf{c} + \alpha^*\mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\hat{\mathbf{q}}_{k+1} \end{aligned} \tag{25}$$

2) Iterate: for $k=1,2,\ldots,N$

$$\begin{aligned} \mathbf{G}_k &= \left[\frac{\mathbf{A}_k^H(\phi)\mathbf{A}_k(\phi)}{\sigma^2} + \mathbf{C}^{-1}\right. \\ &\quad + \left.|\alpha|^2\left(\mathbf{C}^{-1} - \mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\mathbf{C}^{-1} - \mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\mathbf{C}^{-1}\right)\right]^{-1} \\ \mathbf{q}_k &= \frac{\mathbf{A}_k^H(\phi)\mathbf{x}_k}{\sigma^2} + (1-\alpha^*)\mathbf{C}^{-1}\mathbf{c} \\ &\quad + \alpha\mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\hat{\mathbf{q}}_{k-1} + \alpha^*\mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\hat{\mathbf{q}}_{k+1} \\ \mathbf{g}_k &= \mathbf{G}_k\mathbf{q}_k \quad . \end{aligned} \tag{26}$$

$\mathbf{G}_k$ and $\mathbf{g}_k$ from above will be used in $Q(\theta|\phi)$ for the M-step cost function for time $k$.

3) Update for the next EM iteration:

$$\begin{aligned} \hat{\mathbf{G}}_k &= \left[\frac{\mathbf{A}_k^H(\phi)\mathbf{A}_k(\phi)}{\sigma^2} + \mathbf{C}^{-1} + |\alpha|^2\left(\mathbf{C}^{-1} - \mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\mathbf{C}^{-1}\right)\right]^{-1} \\ \hat{\mathbf{q}}_k &= \frac{\mathbf{A}_k^H(\phi)\mathbf{x}_k}{\sigma^2} + (1-\alpha^*)\mathbf{C}^{-1}\mathbf{c} + \alpha\mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\hat{\mathbf{q}}_{k-1}, \end{aligned} \tag{27}$$

with $\hat{\mathbf{G}}_0 \equiv \hat{\mathbf{G}}_{N+1} \equiv \mathbf{C}$, $\hat{\mathbf{q}}_0 \equiv \left(\frac{1-\alpha^*}{1-\alpha}\right) \mathbf{C}^{-1}\mathbf{c}$, and $\hat{\mathbf{q}}_{N+1} \equiv \mathbf{C}^{-1}\mathbf{c}$.

It is apparent from the above algorithm that $f(\mathbf{s}_k|\mathbf{X}, \phi)$ at any time $k$ depends on *all* of the received data and estimated symbols from time 1 to $N$.

The detailed EM algorithm for Gauss-Markov $\mathbf{S}$ is:

1. Initialize $\phi$ to an estimate of $\theta$.
2. E-step: use the algorithm from (25-27) to estimate $\mathbf{G}_k$ and $\mathbf{g}_k$, $k = 1, \ldots, N$.
3. M-step: use the cost function (12) to estimate $\theta$.
4. Set $\phi = \theta$ and repeat steps 2 and 3 until converged.

To initialize $\phi$ in step 1, we could, for example, initialize $\mathbf{G}_k$ and $\mathbf{g}_k$ using $\phi = 0$ in (25-27).

### 4.3. Discussion

An EM iteration is particularly simple, consisting of the calculation of the sufficient satistics $\mathbf{g}_k$ and $\mathbf{G}_k$ in the E-step and the maximization with respect to $\theta$ in the M-step.

Concerning an E-step, for temporally indepedent signal amplitudes, the $\mathbf{g}_k$ and $\mathbf{G}_k$ are computed independently over $k$ using only simple matrix/vector operations and a $D$-dimensional matrix inverse for each $k$ (e.g. for Gaussian signal amplitudes, (20, 21) ). While more computationally complex for the Gauss-Markov case, since the $\mathbf{g}_k$ and $\mathbf{G}_k$ are forward/backward recursive functions of all the data (i.e. (25-27) ), computation of these quantities is still based just on simple matrix/vector operations and $D$-dimensional matrix inverses.

Concerning an M-step, maximization of (12) is comparable in computation to the solution of the CML problem, where initialization is not an issue assuming the EM algorithm has been properly initialized. Computation per M-step iteration can be reduced using Generalized EM (GEM).

Initialization and convergence are important issues for any EM algorithm, or for any ML algorithm for that matter. For each case addressed above we suggest an approach to EM algorithm initialization. Concerning convergence, the required number of E-step/M-step iterations will depend on the application scenario and on the required accuracy. We do not address this important issue in this paper, except to rationalize our selection of the number-of-iterations used for each simulation.

## 5. SIMULATION RESULTS

Here we present computer simulations which show that the EM source location estimation algorithm for Gaussian distributed $f(\mathbf{S})$ (from Section 4.1) performs better than CML, providing performance closer to that of ML estimation with known signal amplitudes. In these simulations two sources, impinging from $14^o$ and $16^o$, were equipowered, mutually uncorrelated, and Gaussian distributed with mean 0. An equispaced linear array of 10 omnidirectional and matched sensors with half wavelength spacing was used. SNR was measured at a sensor. Twenty independent snapshots were employed for each estimate.

To assure results were not skewed by global convergence effects, the known signal ML algorithm was initialized using the true value of $\theta$. To avoid global convergence effects and assure convergence of the iteration process, the EM algorithm was initialized

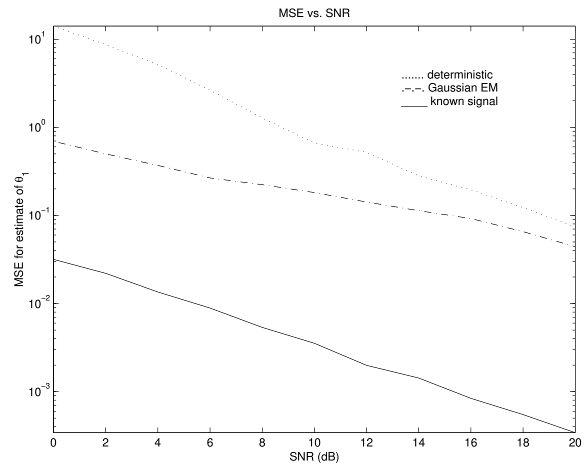


Figure 1: MSE in the estimate of $\theta_1$ vs. SNR for CML, Gaussian EM, and known signal algorithms.

with the known signal ML solution. The CML algorithm, which minimizes $\sum_{k=1}^{N} ||\mathbf{x}_k - \mathbf{A}(\theta)\mathbf{s}_k||^2$ with respect to both $\theta$ and $\mathbf{S}$ using $\mathbf{s}_k = \mathbf{A}(\theta)^+\mathbf{x}_k$, was also initialized with the known signal ML solution.

For the first simulation, 1000 trials were used per point in SNR. For the EM algorithm, the temporally independent Gaussian signal amplitude model was used and the correct signal covariances $\mathbf{C}_k$ and means $\mathbf{d}_k$ were assumed. Enough iterations were used to assure EM algorithm convergence. (For the higher SNR's, the number of iterations was very large - over 100 - although we used a generalized EM algorithm which significantly reduced computation per iteration. In practice, a smaller number of iterations could probably be used.) Figure 1 shows the mean-square-error in the estimate of the $14^o$ source. The performance of the Gaussian EM algorithm is better than the deterministic algorithm, especially at lower SNR. From (20, 21) we see that with zero-mean signal amplitudes and at higher SNR, the $\mathbf{g}_k$ approach the projections of the data $\mathbf{x}_k$ onto the iterating signal subspace. Asymptotically (in SNR) the EM algorithm iteratively computes the solution to the joint location/signal estimation problem.

The second simulation illustrates the effect of using an incorrect prior distribution for the signal amplitudes. Again, the temporally independent Gaussian signal amplitude model, with zero mean, was correctly assumed. The SNR for each of the two sources was fixed at 8dB (i.e. $10^{0.8}$ signal variance). The signal variance for the Gaussian prior distribution was varied from $10^{-2}$ to $10^3$ (the model correctly assumed the sources were mutually uncorrelated and zero mean). 2000 trials were used. Again, enough iterations were used to assure EM algorithm convergence. Figure 2 shows the mean-square-error, in the estimate of the $14^o$ source, for the EM algorithm as a function of model variance. Also shown are the mean-squared-error for known signal ML and CML estimators (which are not a function of model variance). For model variance too small, performance is degraded. This should be expected since the marginalization over $\mathbf{S}$ is then based on a prior that does not represent the actual signal values. When the model variance is larger than that for the actual signal amplitudes, performance also degrades. However, this degradation is not as substantial, since the model still does represent the actual signal values. As the

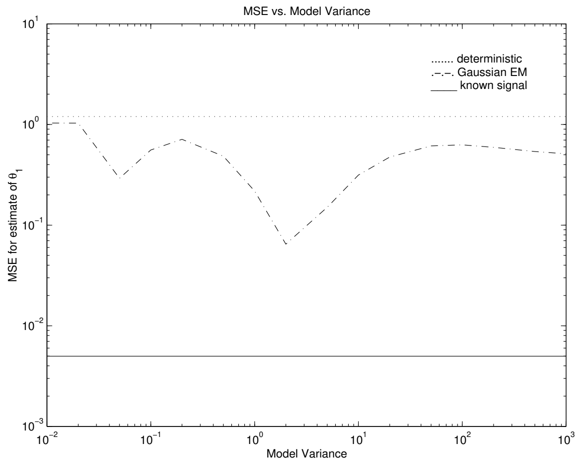


Figure 2: MSE in the estimate of $\theta_1$ vs. SNR for CML, Gaussian EM, and known signal algorithms.

model variance increases, performance asymptotically approaches that realized with a noninformative prior (e.g. [4]). This illustrates the advantage of marginalization even when an accurate prior can not be identified.

## 6. SUMMARY

In this paper we have considered incorporation of knowledge of prior distributions on both primary location and nuisance signal amplitude parameters into the conditional maximum likelihood source location estimation framework. We have focused on using prior distributions on the signal amplitude parameters to marginalize over them. Under the assumption of additive Gaussian noise, we have introduced an EM algorithmic approach to solving the resulting optimization problem, which is in general intractable to solve directly.

We have derived specific EM algorithms for two cases: the temporally independent Gaussian signal amplitude case; and the time dependent signal amplitude case where temporal dependence is modeled as Gauss-Markov. These algorithms are computationally simple in that the E-step computations are straightforward, and the M-step optimization problem is comparable to that of CML source localization. For the temporally independent Gaussian signal amplitude case, we have provided simulations which establish the efficacy of the proposed EM algorithmic approach, illustrate the performance advantage of marginalization compared to joint estimation, and demonstrate the effect of mismatch between the actual and prior nuisance parameter distributions.

## References


1. J.F. Bohme, "Estimation of spectral parameters of correlated signals in wavefields", *Signal Processing*, Vol. 11, pp. 329-337, 1986.
2. Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, pp. 1081-1089, Oct., 1986.
3. P. Stoica and A. Nehorai, "MUSIC, maximum-likelihood, and Cramer-Rao bound", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-37, pp. 720-741, May, 1989.
4. B. Radich and K. Buckley, "Single-snapshot DOA estimation and source number detection", *IEEE Signal Processing Letters*, Vol. 4, No. 4, pp. 109-111, Apr. 1997.
5. A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. Roy. Stat. Soc.*, Vo. 39, No. 1, pp. 1-38, 1997.
6. M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-36, pp. 477-489, Apr., 1988.
7. M. Miller and D. Fuhrmann, "Maximum-likelihood narrowband direction finding with the EM algorithm", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-38, pp. 1560-1577, Sept., 1990.
8. I. Ziskind and D. Hertz, "Maximum-likelihood localization of narrow-band autoregressive sources via the EM algorithm, *IEEE Transactions on Signal Processing*, Vol. ASSP-41, pp. 2719-2724, Aug., 1993.
9. R. Perry, W. Berger and K. Buckley, "EM algorithms for sequence estimation over random ISI channels", *33-rd Asilomar Conf. on Circuits, Systems and Computers*, Oct., 1999.
10. R. Perry, W. Berger and K. Buckley, "EM algorithm for sequence estimation over Gauss-Markov ISI channels", *ICC 2000 Conference*, June 2000.